Genomic Imprinting in the Endosperm Is Systematically Perturbed in Abortive Hybrid Tomato Seeds

Florez-Rueda, Ana M ; Paris, Margot ; Schmidt, Anja ; Widmer, Alex ; Grossniklaus, Ueli ; Städler, Thomas

DOI: https://doi.org/10.1093/molbev/msw175

Genomic Imprinting in the Endosperm is Systematically Perturbed in Abortive Hybrid Tomato Seeds

Ana M. Florez-Rueda,[1,3] Margot Paris,[1,4] Anja Schmidt,[2] Alex Widmer,[1]
Ueli Grossniklaus,[2] and Thomas Städler[*,1]

[1] Plant Ecological Genetics, Institute of Integrative Biology & Zurich–Basel Plant Science Center, ETH Zurich, 8092 Zurich, Switzerland

[2] Plant Developmental Genetics, Department of Plant and Microbial Biology & Zurich–Basel Plant Science Center, University of Zurich, 8008 Zurich, Switzerland

[3] Present address: see [2]

[4] Present address: Unit of Ecology & Evolution, Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland

[*] **Corresponding author**: Thomas Städler, thomas.staedler@env.ethz.ch

Keywords: endosperm, seed failure, postzygotic isolation, genomic imprinting, *Solanum*

Short Title: Expression perturbation in hybrid endosperm

Abstract

Hybrid seed failure represents an important postzygotic barrier to interbreeding among species of wild tomatoes (*Solanum* section *Lycopersicon*) and other flowering plants. We studied genome-wide changes associated with hybrid seed abortion in the closely related *Solanum peruvianum* and *S. chilense* where hybrid crosses yield high proportions of inviable seeds due to endosperm failure and arrested embryo development. Based on differences of seed size in reciprocal hybrid crosses and developmental evidence implicating endosperm failure, we hypothesized that perturbed genomic imprinting is involved in this strong postzygotic barrier. Consequently, we surveyed the transcriptomes of developing endosperms from intra- and interspecific crosses using tissues isolated by laser-assisted microdissection. We implemented a novel approach to estimate parent-of-origin–specific expression using both homozygous and heterozygous nucleotide differences between parental individuals and identified candidate imprinted genes. Importantly, we uncovered systematic shifts of 'normal' (intraspecific) maternal:paternal transcript proportions in hybrid endosperms; the average maternal proportion of gene expression increased in both crossing directions but was strongly negatively correlated with 'normal' maternal proportions. These genome-wide shifts almost entirely eliminated paternally expressed imprinted genes in *S. peruvianum* hybrid endosperm but also affected maternally expressed imprinted genes and all other assessed genes. These profound, systematic changes in parental expression proportions suggest that core processes of transcriptional regulation are functionally compromised in hybrid endosperm and contribute to hybrid seed failure.

Introduction

Elucidating the evolutionary processes underlying the establishment of reproductive isolation between recently diverged lineages, as well as its molecular underpinnings, remains a fundamental problem in explaining the origins of biodiversity (Coyne and Orr 2004; Crespi and Nosil 2013). A somewhat neglected phenomenon in this regard is hybrid seed failure in flowering plants, albeit early work recognizing it as widespread and of significance for the potentially rapid establishment of postzygotic reproductive isolation (Cooper and Brink 1940; Brink and Cooper 1947; Haig and Westoby 1991; Lester and Kang 1998; Bushell et al. 2003). Following successful double fertilization, growth of the endosperm—an essential seed compartment in angiosperms—typically shows aberrant features in such crosses and eventually results in embryo and seed abortion.

Many empirical observations point to a decisive role for parental genome dosage and sensitivity to such dosage in the success or failure of particular crosses. For example, interploidy crosses typically have large effects on endosperm size (Cooper and Brink 1945; Lin 1984; Birchler 1993; Scott et al. 1998), and failure of crosses between different ploidy levels within species sometimes resembles failure in interspecific crosses of the same ploidy (Bushell et al. 2003; Gutierrez-Marcos et al. 2003). These concordant observations have fueled the hypothesis that parent-of-origin–dependent gene expression (genomic imprinting) might be causally involved in hybrid seed failure (Haig and Westoby 1991; Gutierrez-Marcos et al. 2003). Genomic imprinting is an epigenetic phenomenon known in angiosperms and mammals and refers to Allele-Specific Expression (ASE) that depends on whether the allele was inherited from the female or the male parent. In flowering plants, imprinting is most prevalent in the (normally) triploid endosperm and is critical for its proper development and thus for normal seed development (Raissig et al. 2011; Jiang and Köhler 2012; Gehring 2013).

In a seminal paper, Haig and Westoby (1991) interpreted the effects of between-species and interploidy crosses on seed development as reflecting genetic conflicts between maternally and paternally derived alleles over the allocation of resources from mother to offspring (also known as the ‘kinship theory’ or ‘parental conflict theory’ for the evolution of imprinting; Haig 2013; Pires and Grossniklaus 2014). Under this model, imprinting and the resulting levels of gene expression collectively secure successful seed development in the context of antagonistic parental forces. Normal seed development is therefore expected to be sensitive to changes in ploidy or any molecular divergence between parents that would affect genomic imprinting. However, the underlying cause of (hybrid) seed failure must not

necessarily be sought in perturbed imprinting. More recently, alternative molecular mechanisms—that might well act in concert with perturbed imprinting—have been proposed to account for seed failure, such as small interfering (si) RNAs and the derepression of Transposable Elements (TEs) mediated by siRNAs (Castillo and Moyle 2012; Lu et al. 2012; Ng et al. 2012; Lafon-Placette and Köhler 2015).

With the advent of Next-Generation-Sequencing technologies, it has become possible to assess entire endosperm transcriptomes for evidence of genomic imprinting. Consequently, hundreds of candidate imprinted genes have been identified in *Arabidopsis thaliana* (Gehring et al. 2011; Hsieh et al. 2011; Wolff et al. 2011; Pignatta et al. 2014), rice (Luo et al. 2011; Rodrigues et al. 2013), and maize (Zhang et al. 2011; Waters et al. 2013). One of the major recent discoveries concerns evidence for allele-specific imprinting and, more generally, variation within species for imprinting status (Waters et al. 2013; Pignatta et al. 2014). These observations were mechanistically explained by transiently altered methylation patterns downstream of evolutionarily recent TE insertions (Pignatta et al. 2014), and imply that only a small minority of imprinted genes may be functionally important for normal endosperm and seed development. However, these recent endosperm RNA-Seq studies on plant model systems did not focus on reproductive isolation and the potential involvement of perturbed or dissimilar patterns of imprinting.

Previous investigations of the molecular signatures of hybrid seed failure are restricted to the genus *Arabidopsis*. Arguably the best-studied examples of interspecific hybrid seed failure at the same ploidy level involve *A*. *thaliana* × *A*. *arenosa* crosses (Josefsson et al. 2006; Walia et al. 2009; Burkart-Waco et al. 2013, 2015). This body of work documented biparental expression patterns of normally imprinted genes (*MEDEA* and *PHERES1*) in hybrid endosperm, as well as the reactivation of retrotransposons. Josefsson et al. (2006) established a causal link between perturbation of imprinting and the degree of interspecific seed abortion, equivalent to similar work on interploidy crosses in *A*. *thaliana* (Jullien and Berger 2010; Kradolfer et al. 2013). Moreover, these studies imply that this type of postzygotic barrier could be erected by evolutionary changes at very few genes. The most recent study of *A*. *thaliana* × *A*. *arenosa* crosses (Burkart-Waco et al. 2015) found evidence for a general shift toward higher maternal transcript proportions in hybrid endosperm and the concomitant mis-expression of paternally expressed imprinted genes. Because entire seeds were the source of RNA, however, this inference had to be restricted to the subset of genes expressed exclusively in the *A*. *thaliana* endosperm.

Wild tomatoes (*Solanum* section *Lycopersicon*) comprise close relatives of the cultivated tomato and exhibit variable levels of postzygotic isolation among pairs of taxa (Rick and Lamm 1955; Rick 1979, 1986; Peralta et al. 2008). In classical studies, C.M. Rick found very high proportions of hybrid seed failure in reciprocal crosses of *Solanum peruvianum* and *S. chilense* (Rick and Lamm 1955; Rick 1979, 1986), two closely related species with partly overlapping ranges in northern Chile and southwestern Peru. A few F1 hybrid seeds 'escaped' abortion and after germination proved to be viable hybrid plants (Rick and Lamm 1955; Rick 1986), suggesting that the normal failure of such seeds can be attributed to disturbed endosperm–embryo interactions rather than early-acting incompatibilities between the two parental genomes in F1 embryos. This interpretation is strengthened by the success of F1 embryo culture in several interspecific crosses in the tomato clade, i.e. aborting embryos can be rescued by excising them from the seed and culturing them *in vitro* (Brink and Cooper 1947; Rick and Lamm 1955; Rick 1979).

Motivated by their evolutionarily interesting suite of biological properties, *S. peruvianum* and *S. chilense* have been the object of recent multilocus studies focusing on demography and speciation history, molecular evolution, and abiotic adaptation (e.g. Städler et al. 2005, 2008; Tellier et al. 2011; Böndel et al. 2015). Here, we provide evidence for genomic imprinting based on reciprocal crosses within both *Solanum* species, using endosperm tissue isolated by laser-assisted microdissection (LAM) as the source of RNA. Importantly, we characterized ASE in failing endosperm from the reciprocal hybrid crosses. While our work should not be considered a comprehensive study of imprinting in these taxa due to possible intraspecific variation in imprinting (Waters et al. 2013; Pignatta et al. 2014), it provides the first near-unbiased perspective on genome-wide changes in maternal:paternal transcript proportions in failing hybrid endosperm in flowering plants.

Results and Discussion

Phenotypic Asymmetry of Inviable Hybrid Seeds

Interspecific crosses between the two wild tomato accessions, one representing *S. peruvianum* and one *S. chilense*, respectively, resulted in almost complete seed failure, as expected from earlier studies (Rick and Lamm 1955; Rick 1979). Consequently, we recovered almost no viable seeds in interspecific crosses using the populations chosen for molecular analyses, in strong contrast to within-population crosses that yielded high proportions of viable seeds (fig. 1*A*). Importantly, the number of seeds per fruit was not significantly different between any of

the six cross-type comparisons (Wilcoxon rank-sum tests, all $p > 0.05$), emphasizing that there was no discernable post-mating, prezygotic interspecific barrier under the imposed noncompetitive pollination conditions (fig. 1*A*).

While almost exclusively yielding inviable seeds, the reciprocal interspecific crosses exhibited distinctly different seed size, with hybrid seeds being markedly smaller on *S. peruvianum* maternal plants (fig. 1*B*). Highly stable differences in average seed size have previously been documented for reciprocal interspecific and interploidy crosses (Scott et al. 1998; Lu et al. 2012; Willi 2013). Such differences have widely been attributed to parental genome conflict (Haig and Westoby 1991; Brandvain and Haig 2005), with relatively larger seeds exhibiting a 'paternal excess' phenotype (thought to be due to enhanced nutrient-acquiring ability of the endosperm) in contrast to relatively smaller seeds exhibiting a 'maternal excess' phenotype (thought to reflect balanced distribution of maternal resources amongst all seeds). The characteristic maternal excess phenotype of hybrid seeds produced by the *S. peruvianum* × *S. chilense* cross (fig. 1*B*) may be functionally linked to the larger increase in the maternal proportion of endosperm transcripts as revealed by our ASE analyses (see below).

Genomic Imprinting in the Endosperm

Two sets of reciprocal crosses were conducted within species to assess genomic imprinting in the 'normal' endosperm of wild tomatoes. Deep sequencing of RNA obtained from endosperms isolated by LAM at 14 Days After Pollination (DAP) yielded a large number of sequencing reads (48–74 million across two replicate libraries) for each of the four genotypes. An average of 83.7% of the reads could be uniquely mapped to the gold-standard tomato reference genome (The Tomato Genome Consortium 2012; for details see supplementary table S1, Supplementary Material online). In contrast to *A. thaliana*, our focal plants are obligate outcrossers and both species harbor fairly high levels of nucleotide diversity (Städler et al. 2005, 2008; Tellier et al. 2011; Böndel et al. 2015); thus, we expected the majority of sequence differences between the parents to occur in a heterozygous state. To make use of information both from homo- and heterozygous parental differences, we implemented a novel approach to integrate differences between the parents of the type CC:AC (i.e. where one parent is a homozygote and the other a heterozygote carrying another nucleotide; for details see *Materials and Methods* and supplementary fig. S1, Supplementary Material online). All sites with a minimum coverage of 10 reads were used to obtain a transcript-specific estimate

of the maternal expression proportion. After filtering, ASE was successfully estimated for a total of 8,229 genes in *S. peruvianum* LA1616 and 2,560 genes in *S. chilense* LA4329, reflecting the higher number of parental sequence polymorphisms in the S. *peruvianum* cross (table 1).

The endosperm has a genomic composition of 2:1 maternal:paternal haploid genomes, such that we expect the 'normal' proportion of maternal expression to be close to 0.667. Our empirical data broadly reflect these expectations, but in line with equivalent data from other plant studies (Waters et al. 2013; Pignatta et al. 2014), there was a lot of scatter in the distribution of maternal proportion estimates across individual transcripts (fig. 2*A* and supplementary fig. S2*A*, Supplementary Material online). Candidate imprinted Maternally Expressed Genes (MEGs) exceed our imposed threshold of 0.833 maternal proportion in both cross directions, and candidate imprinted Paternally Expressed Genes (PEGs) are those with <0.333 maternal proportion in both cross directions. For *S. peruvianum* LA1616, we identified 351 candidate MEGs (fig. 2*A*, upper right, colored box) and 172 candidate PEGs (fig. 2*A*, lower left, colored box). The corresponding numbers for the less informative *S. chilense* LA4329 are 40 candidate MEGs and 70 candidate PEGs (table 1 and supplementary fig. S2*A*, Supplementary Material online). Details regarding Single Nucleotide Polymorphisms (SNPs), maternal proportions and functional annotation of candidate MEGs and PEGs identified in each species can be found in supplementary table S2, Supplementary Material online.

Imprinted Genes' Functional Roles and their Evolutionary Maintenance

Given our main focus on hybrid seed failure and its molecular correlates, here we highlight only novel aspects not predictable from previous large-scale endosperm RNA-Seq studies in plant model systems (Gehring et al. 2011; Hsieh et al. 2011; Luo et al. 2011; Wolff et al. 2011; Zhang et al. 2011; Rodrigues et al. 2013; Waters et al. 2013; Pignatta et al. 2014). Of note, a total of 35 genes potentially encoding subunits of Skp1–Cullin–F-box (SCF) protein complexes are among our candidate imprinted genes (9 MEGs and 26 PEGs; blue highlight in supplementary table S2, Supplementary Material online). Functioning of the SCF complex relies on proper coupling of its core proteins and cofactors. If genes coding for components of the SCF complex acquired imprinted expression for their role in modulating seed development as hypothesized by Dumbliauskas et al. (2011), the imprinted expression of such a high number of genes in wild tomatoes may have evolved as a result of coadaptation of gene expression, as posited by Wolf's (2013) model. Second, 30 nuclear-encoded chloroplast genes

were found to be maternally expressed (green highlight in supplementary table S2, Supplementary Material online). Finding such a high number of nuclear-encoded genes whose protein products work in concert with chloroplast-encoded subunits as candidate MEGs is unprecedented in previous studies of imprinting in angiosperms (see references above). Interestingly, these results fit expectations under Wolf's (2009) cytonuclear interactions model where nuclear-encoded organelle genes evolved to be maternally imprinted owing to coadaptation with organelle metabolism. Thirty of the 32 candidate imprinted nuclear-encoded chloroplast genes in wild tomatoes are MEGs, consistent with this model for imprinting due to cytonuclear epistasis of nuclear and chloroplast genomes (Wolf 2009).

Genome-Wide Increase of Maternal Transcript Proportions in Hybrid Endosperms

The reciprocal hybrid cross yielded 4,111 transcripts that could be assessed for their maternal:paternal expression proportions (table 1). Intriguingly, overall maternal transcript proportions were elevated in both directions of the hybrid cross (fig. 2*B*). However, this trend was more pronounced for hybrid seeds developing on the *S. peruvianum* 1616A plant, exhibiting an increase in median maternal proportion from 0.646 in the within-population cross to 0.843 in the hybrid cross. The shift in median maternal transcript proportion on the *S. chilense* 4329B plant was more modest, from 0.630 in the within-population cross to 0.701 in the hybrid cross (table 1).

To explore the scale and direction of the changes at the gene level, we calculated the magnitude of the shift in maternal proportion from normal endosperm to hybrid endosperm (Δ in maternal proportion hyb–sib) for all genes with parental sequence differences in both within-population and hybrid crosses (*S. peruvianum*: 3,647 genes; *S. chilense*: 1,856 genes). The differences in maternal proportions between hybrid and normal endosperm plotted against the maternal proportions derived from within-population endosperms on the same maternal plants are shown in fig. 3*A* and supplementary fig. S2*B*, Supplementary Material online. We found striking negative correlations between Δ in maternal proportion hyb–sib and maternal proportions in within-population crosses in the data of both species (Spearman's rank correlation, $\rho = -0.774$ for *S. peruvianum* and –0.652 for *S. chilense*; both $p < 10^{-15}$). In other words, the lower the maternal contribution to expression in normal endosperm, the larger the maternal contribution in hybrid endosperm. Surprisingly, while the general pattern in hybrid endosperms shows elevated maternal transcript proportions, the strong negative correlations identified above entail the opposite shift at high maternal proportions in the within-population

data. These genes (among them many candidate MEGs) do not show an average increase in maternal proportions but tend to exhibit a slight decrease of maternal transcript proportions in hybrid endosperms (fig. 3*A* and supplementary fig. S2*B*, Supplementary Material online).

Candidate imprinted genes do not seem to 'escape' these broad patterns, although as a group, PEGs experience a smaller shift in maternal proportion in hybrid endosperm compared to the non-imprinted genes from the same range of low maternal proportions (i.e. <0.333; fig. 3*A* and supplementary fig. S2*B* [left side], Supplementary Material online). The difference in maternal proportion is significantly smaller for candidate PEGs in both species (*S. chilense*, median Δ PEGs = 0.276, median Δ non-PEGs = 0.466, Wilcoxon rank-sum test p <0.001; *S. peruvianum*, median Δ PEGs = 0.532, median Δ non-PEGs = 0.598, Wilcoxon rank-sum test p <0.002). For the high range of maternal transcript proportions (i.e. >0.833), non-MEGs experience a significantly larger difference in maternal proportion than candidate imprinted MEGs in one of the species (*S. chilense*, median Δ MEGs = –0.036, median Δ non-MEGs = –0.166, Wilcoxon rank-sum test p = 0.024; *S. peruvianum*, median Δ MEGs = –0.018, median Δ non-MEGs = –0.026, Wilcoxon rank-sum test p = 0.36). These results suggest that, at PEG loci, maternal alleles that are normally repressed become de-repressed in hybrid tomato endosperm. However, as this perturbation of normal transcriptional regulation is not confined to imprinted genes but affects the entire spectrum of maternal expression proportions, we conclude from these patterns that the regulatory machinery of transcription is fundamentally compromised in hybrid endosperm.

Perturbation of Candidate MEGs and PEGs in Hybrid Endosperms

We further quantified changes in the parental transcript proportions of candidate imprinted genes, separately for both species. As a consequence (or more precisely, an epiphenomenon) of the genome-wide shift toward higher hybrid maternal transcript proportions, the paternal expression bias of many candidate PEGs was eliminated in hybrid endosperm (supplementary figs. S3*A*, *C*, Supplementary Material online). Furthermore, this shift toward higher maternal proportions in candidate PEGs is larger in the *S. peruvianum* LA1616 data (median Δ hyb–sib = 0.532) than in the *S. chilense* LA4329 data (median Δ hyb–sib = 0.276), consistent with the global trend of stronger maternal allelic bias in the hybrid endosperm of *S. peruvianum* (table 1; fig. 3*A*). About 47% (18/38) of the *S. chilense* candidate PEGs remain PEGs in the hybrid cross with 1616A as the male parent (supplementary fig. S3*C* [shift along the *x* axis], Supplementary Material online). In keeping with the generally smaller shifts in the 'high'

range of maternal proportion, most candidate MEGs remain MEGs in hybrid endosperm of both species (86% in *S. peruvianum* [124/145] and 62% in *S. chilense* [13/21]; supplementary figs. S3*B*, *D*, Supplementary Material online).

Finally, we evaluated differences in maternal transcript proportions between within-population and hybrid endosperms for the smaller group of genes with evidence for being imprinted in both *Solanum* species (so-called 'shared' imprinted genes; fig. 3*B*). The average shift for candidate PEGs is less severe for *S. chilense* in the maternal role in this set of genes, but the overall pattern mirrors that of the larger set of species-specific candidate PEGs (supplementary fig. S3, Supplementary Material online), in particular the sweeping disruption of paternal expression bias in *S. peruvianum*. Likewise, the shared MEGs show rather slight departures from their corresponding estimates in intra-population endosperms, with most of them retaining their MEG status in hybrid endosperms of both species (fig. 3*B*). This pattern is consistent with that in the larger set of species-specific candidate MEGs (supplementary fig. S3, Supplementary Material online).

Recent work on hybridization between *A. thaliana* and *A. arenosa* has uncovered similar perturbations in maternal transcript proportions for specific genes, with several previously known PEGs exhibiting normal-to-high expression from the maternal allele in hybrid seeds, among other changes (Burkart-Waco et al. 2015). These similarities may reflect shared responses to hybridization in both Brassicaceae and Solanaceae, possibly due to equivalent molecular mechanisms involved in the establishment of imprinted expression as well as its misregulation upon interspecific hybridization. Burkart-Wako et al. (2015) argued that *Polycomb* Repressive Complex 2 may be responsible for this behaviour, due to its known role in regulating some imprinted loci in *Arabidopsis* (e.g. Köhler et al. 2005; Jullien et al. 2006). Another transcriptional regulatory mechanism in the developing seed is RNA-directed DNA methylation (RdDM), which regulates imprinting at specific loci expressed in the endosperm and is mediated by siRNAs (Vu et al. 2013). While RdDM has been shown to regulate imprinting at a handful of *Arabidopsis* loci, we consider this mechanism improbable to account for the transcriptome-wide trend we have uncovered; such a scenario would imply that all genes undergoing shifts in maternal expression in hybrid endosperm have TEs in their respective genomic neighborhoods that are targeted by RdDM. Alternatively, the RdDM pathway may also regulate non-TE targets in the endosperm as it was found in young *Arabidopsis* embryos, where the lack of maternal components of the RdDM pathway leads to de-repression of paternal alleles at many loci (Autran et al. 2011). Given the general, genome-

wide shift of maternal transcript proportions, however, it is more likely that general functions in transcription are affected. We speculate that the composition of multimeric complexes involved in transcription and its regulation is optimized for its respective genome. In hybrid endosperms, more of these subunits are encoded by the maternal genome, leading to the formation of a higher proportion of multimeric complexes that are composed of mostly maternal isoforms, thus transcribing the maternal genome more efficiently.

Conclusions

Our interspecific crosses have uncovered unprecedented, systematic shifts of maternal transcript proportions in (failing) hybrid endosperms of two wild tomato species. Related work on *Arabidopsis* was constrained by targeting known candidate imprinted genes, particularly PEGs (Burkart-Waco et al. 2015; Wolff et al. 2015), and could not assess the breadth of transcriptional changes because RNA was isolated from entire seeds in these studies. The shifts in maternal transcript proportions documented here clearly affect a majority of endosperm-expressed genes and are not restricted to imprinted genes, nor solely PEGs or MEGs. Interestingly, our data reveal that candidate PEGs are *less* affected than other genes in the same range of maternal expression proportions, yet their average shift toward higher maternal proportion is sufficient to eliminate their 'normal' paternal expression bias in the hybrid yielding very small seeds (*S. peruvianum* as female parent). These intriguing patterns notwithstanding, which of the assessed changed expression properties (if any) in hybrid endosperm contribute to interspecific seed failure in these wild tomatoes cannot be determined with the data at hand. Future studies will encompass more crosses with greater numbers of parental nucleotide differences, evaluating expression-level changes, and work with a more mechanistic focus to investigate the molecular basis of the genome-wide shift in maternal transcript proportions.

Materials and Methods

Plant Material, Crossings and Seed Evaluation

All seeds were obtained from the C.M. Rick Tomato Genetics Resource Center at U.C. Davis (http://tgrc.ucdavis.edu). Representing *Solanum peruvianum*, we used seeds from accession LA1616 (Lima, Peru), and representing *S. chilense*, we used seeds from accession LA4329 (Antofagasta, Chile). Both accessions are strictly self-incompatible, eliminating the need to emasculate flowers before applying pollen from a different plant. Several plants per accession were grown under standard, insect-free greenhouse conditions. Freshly opened flowers were used both as sources and recipients of pollen from other plants, either from the same accession (that is, within-LA1616 and within-LA4329 crosses) or among accessions (that is, heterospecific crosses). Pollinations were performed by manually collecting pollen from the paternal plant and immediately transferring it to the stigmas of the maternal plant. Stigmas were completely covered with pollen to secure enough seed production. All hand-pollinated flowers were individually marked and ripe fruits were collected 60 DAP. All resulting seeds (viable and non-viable) were counted and seed viability was determined by the presence of a fully developed embryo that had reached a coiled stage, irrespective of seed size. Statistical analyses and plotting were performed using the R software, version 3.2.1 (R Development Core Team 2014).

Crossing Design for Endosperm Transcriptome Sequencing

For the molecular component of this study, we used four of the plants from among the larger cohort, referred to as 1616A, 1616J, 4329B and 4329K. We analyzed three different parental combinations: the within-species *S. peruvianum* case with plants 1616A and 1616J as parents, the within-species *S. chilense* case with plants 4329B and 4329K as parents, and the between-species case with plants 1616A and 4329B as parents. For each of these four parental plants, transcriptomes were obtained by sequencing RNA from flower buds. Young flower bud tissue was collected in liquid nitrogen and RNA extracted with the RNAeasy mini RNA isolation kit (Qiagen), according to the manufacturer's instructions. Libraries were prepared with the Illumina TrueSeq RNA Sample Preparation Kit v2 following the manufacturer's instructions and were sequenced on the Illumina HiSeq2000 platform, generating 150-bp paired-end reads.

We chose to base our exploratory work on genomic imprinting on intra-population crosses (at the possible cost of having lower power to distinguish maternal from paternal reads due to fewer SNPs between parents) to minimize the incidence of failed seed development

observed in many intraspecific, inter-population crosses (our unpublished data). Several months before the beginning of the experiment, the four focal plants were transferred from the greenhouse to climate-controlled chambers. The conditions in the climate chambers were 12 hours of light (18 Klux) at 22°C with 50% relative humidity and 12 hours of darkness (0 Klux) at 18°C with 60% relative humidity. For each of the three cross types, reciprocal hand pollinations were performed and developing fruits were collected on each plant for each cross type. Based on previous observations on seed development in *Solanum* (Beamish 1955; Dnyansagar and Cooper 1960; Pacini and Sarfatti 1978; Briggs 1993) and our histological analyses of seed development (unpublished), we chose an early globular embryo stage to collect the material for library preparation. We collected fruits 14 DAP, always in late afternoon. This time point was chosen because it was early enough to distinguish the developing embryo from the surrounding endosperm tissue, while the latter was large enough to extract RNA in the quantities needed for library preparation. For each plant and cross type, two separate RNA libraries were prepared from laser-captured endosperm tissue, for a total of 12 endosperm libraries.

Laser Microdissection, RNA Extraction and Library Preparation

For analysis of the endosperm transcriptomes, the collected fruits were immediately placed in a fixation solution (9:1 ethanol:acetic acid). All solutions were maintained cold at <4°C from fixation to transfer of the samples into the embedding machine (see below). The samples were swiftly transferred to a refrigerator and remained in the fixing solution for a minimum of 24h and a maximum of 48h. During this fixing step, the samples were submitted to a vacuum for at least 30 minutes to allow infiltration of the fixative. The samples were then transferred to a cold (<4°C) solution of 90% ethanol and shortly thereafter placed on a Leica embedding machine for paraffin infiltration (settings as in Wuest et al. 2010). Prior to LAM, 8-µm sections were prepared from the samples embedded in paraffin blocks with a RM2145 Leica microtome (Leica Microsystems GmbH, Wetzlar, Germany). The sections were mounted on nuclease-free membrane metal-frame slides (MicroDissect GmbH, Wetzlar, Germany) using water. Slides were dried on a heating table at 42°C for a maximum of two hours. The samples were deparaffinized in xylol at room temperature in two fifteen-minute washes.

To isolate endosperm tissue for RNA sequencing we followed the protocol described in Schmid et al. (2012). In brief, LAM was performed using a CellCut Plus device (MMI Molecular Machines & Industries AG, Glattbrugg, Switzerland), carefully separating the

endosperm from the embryo and surrounding sporophytic (i.e. maternal) seed coat tissues. Endosperm tissue was collected using MMI isolation caps and RNA extraction was performed immediately or within 24 hours; in the latter case, the caps were stored at –80°C prior to extraction. RNA was extracted using the Applied Biosystems® Arcturus®PicoPure® RNA Isolation Kit (ref. KIT0204) according to the manufacturer's instructions. The quality and quantity of total RNA was assessed with Agilent Bioanalyzer Pico Chips. RNA that showed clear ribosomal peaks was used for library preparation. We used a minimum of 82 ng RNA for preparing each library, an amount reached by pooling separate extracts. For each library, 200–700 sections of endosperm were isolated, representing several fruits and at least two independent pollination events. Sequencing libraries were prepared with the Illumina TruSeq RNA Sample Preparation Kits v2 following the manufacturer's instructions.

Sequencing and Read Mapping

The 12 endosperm-derived libraries were sequenced on three lanes of the Illumina HiSeq2000 platform at the Functional Genomics Center Zurich (www.fgcz.ch), generating 100-bp single-end reads. The quality of each library was assessed using the FastQC program (http://bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters were removed from the reads with the Cutadapt program (Martin 2011) and quality filtering was performed with the ConDeTri program (Smeds and Künstner 2011) using a minimum quality threshold of 20. RNA-Seq quality-filtered reads for each library were mapped using TopHat version 1.4.0 (Trapnell et al. 2010). Mapping was done against the International Tomato Annotation Group (ITAG) Release 2.3 of the tomato reference genome sequence (The Tomato Genome Consortium 2012) on the SL2.40 genome build, downloaded via the SOL Genomics Network ftp site (ftp://ftp.sgn.cornell.edu/genomes/). A maximum of six mismatches was allowed between reads and the reference, and reads that mapped to more than one position in the genome were discarded.

SNP Calling and ASE Analyses

The mpileup command of SAMtools (Li et al. 2009) was used to call variant sites in the flower bud and endosperm transcriptomes. Popoolation2 (Kofler et al. 2011) was subsequently used to recover the allelic counts of major and minor alleles at each site, using a minimum read mapping quality threshold of 20. Using the allelic counts, we estimated the maternal:paternal transcript proportions separately for each SNP. Our approach is explained below and was implemented in Python 2.7 (scripts are available at https://github.com/anaflo/tomato).

Variant sites between the parental plants were recovered using the flower bud transcriptome sequencing. To make use of as many genotypic differences between the parents as possible, we developed a novel approach integrating information from both heterozygous and homozygous SNPs (table 1 and supplementary fig. S1, Supplementary Material online). The variant sites that can provide information for quantifying ASE are 1) the homozygote cases, that is, reciprocally different homozygous parental genotypes, and 2) the heterozygote cases, i.e. sites at which one parent is homozygous for a given base and the other parent is heterozygous.

Using our custom Python program, we calculated the proportion of the minor allele relative to the major allele for each polymorphic site and determined homozygous and heterozygous SNPs. First, sites with minor allelic proportion <5% were considered as homozygotes in order to avoid biases due to sequencing errors observed in next-generation sequencing. Only reciprocal homozygotes (i.e. different nucleotides between the parental individuals) were kept for the analyses. Then, heterozygote sites were defined using a conservative minor allelic proportion >40% in the heterozygous parent (supplementary fig. S1*A*, Supplementary Material online). For the heterozygous genotypes, the bases were categorized as 'discriminant' or 'fixed'. The 'fixed' base corresponds to the base identity of the homozygous parent, whereas a 'discriminant' base refers to the other base observed in the heterozygous parent. It is the discriminant base that allows the estimation of the endosperm ASE in the heterozygote cases (see below).

In the endosperm data, maternal proportion of overall transcription was calculated based on the parental identity of the alternative bases at each site, i.e. whether a given base in the endosperm data was of maternal or paternal origin (supplementary fig. S1*B*, Supplementary Material online). We note that each of the three reciprocal crosses has independent variant sites dependent on each parental genotype. The two replicated libraries per cross direction were pooled and only sites covered by a minimum of 10 reads were kept for analyses. For reciprocal homozygous sites in the parents, maternal proportion was calculated as the proportion of reads with the maternal base compared to the total number of reads for that site (supplementary fig. S1*B*, Supplementary Material online). For the heterozygous SNPs, we inferred the maternal proportion based on the discriminant base counts, as follows. For paternally heterozygous sites, maternal proportion (mat. prop.) was estimated by subtracting twice the observed proportion of the discriminant base (freq. discr.) from unity, i.e. mat. prop. = 1 – (2 * freq. discr.). Conversely, for maternally heterozygous sites, maternal proportion was

estimated by doubling of the observed discriminant base frequency, i.e. mat. prop. = 2 * freq. discr. (supplementary fig. S1*B*, Supplementary Material online). This rationale assumes 'fair' segregation of alleles at meiosis. Estimates of maternal proportions were constrained to fall within the range 0–1. For genes with multiple informative sites in a given cross, per-gene estimates of maternal proportions were calculated using a weighted average of the independent per-base estimates within genes, implying more weight for more highly covered polymorphic sites within a given gene (supplementary fig. S1*B*, Supplementary Material online).

Once a per-gene value was obtained for each direction of the cross, we used thresholds for maternal proportions in reciprocal crosses to call a given gene as potentially imprinted and consider moderately and strongly imprinted genes, as follows: moderate MEGs >0.833, strong MEGs >0.917, moderate PEGs <0.333 and strong PEGs <0.167 maternal proportion (supplementary fig. S1*B*, Supplementary Material online). These thresholds are reciprocally symmetric and consistent with the expectation of a 0.667 maternal proportion of gene expression in the triploid endosperm (2m:1p). Our 'moderately' and 'strongly' imprinted genes reflect greater than two-fold and four-fold deviations, respectively, from the expected maternal:paternal proportions. Furthermore, genes considered as candidate MEGs and PEGs exhibited significant departures from the expected 2m:1p ratio, as assessed by χ^2 tests with False Discovery Rate corrections. While the above thresholds largely determine whether or not a given gene is considered as candidate imprinted, they have no bearing on the major finding of this study, i.e. the systematic shifts of maternal:paternal transcript proportions in hybrid endosperms (figs. 2, 3, and supplementary figs. S2, S3, Supplementary Material online).

Supplementary Material

Supplementary tables S1-S2, and figures S1-S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Maja Frei for meticulous plant care, Niklaus Zemp for advice on statistics, Mathias Scharmann for help with the Python script and photography for Figure 1*B*, and the Genetic Diversity Center Zurich (GDC) for use of its facilities. We also thank the C.M. Rick Tomato Genetics Resource Center at U.C. Davis for generously supplying seed samples. This work was supported by the Swiss National Science Foundation (31003A_130702 to T.S.), the "Staatssekretariat für Bildung und Forschung" (COST action FA0903 to A.S. and U.G.), funds from the University of Zurich (U.G.), and an ETH Research Grant (ETH-40 13-2 to T.S. and A.W.). The RNA-Seq reads have been deposited at the Sequence Read Archive under accession no. XXXXXXX.

References

Autran D, Baroux C, Raissig MT, Lenormand T, Wittig M, Grob S, Steimer A, Barann M, Klostermeier UC, Leblanc O, et al. 2011. Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* 145:707–719.

Beamish KI. 1955. Seed failure following hybridization between the hexaploid *Solanum demissum* and four diploid *Solanum* species. *Am J Bot*. 42:297–304.

Birchler JA. 1993. Dosage analysis of maize endosperm development. *Annu Rev Genet*. 27:181–204.

Böndel KB, Lainer H, Nosenko T, Mboup M, Tellier A, Stephan W. 2015. North–south colonization associated with local adaptation of the wild tomato species *Solanum chilense*. *Mol Biol Evol*. 32:2932–2943.

Brandvain Y, Haig D. 2005. Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *Am Nat*. 166:330–338.

Briggs CL. 1993. Endosperm development in *Solanum nigrum* L. Formation of the zone of separation and secretion. *Ann Bot*. 72:303–313.

Brink RA, Cooper DC. 1947. The endosperm in seed development. *Bot Rev*. 13:423–541.

Burkart-Waco D, Ngo K, Dilkes B, Josefsson C, Comai L. 2013. Early disruption of maternal-zygotic interaction and activation of defense-like responses in *Arabidopsis* interspecific crosses. *Plant Cell* 25:2037–2055.

Burkart-Waco D, Ngo K, Lieberman M, Comai L. 2015. Perturbation of parentally biased gene expression during interspecific hybridization. *PLoS ONE* 10:e0117293.

Bushell C, Spielman M, Scott RJ. 2003. The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *Plant Cell* 15:1430–1442.

Castillo DM, Moyle LC. 2012. Evolutionary implications of mechanistic models of TE-mediated hybrid incompatibility. *Intl J Evol Biol*. 2012:article ID 698198.

Cooper DC, Brink RA. 1940. Somatoplastic sterility as a cause of seed failure after interspecific hybridization. *Genetics* 25:593–617.

Cooper DC, Brink RA. 1945. Seed collapse following matings between diploid and tetraploid races of *Lycopersicon pimpinellifolium*. *Genetics* 30:376–410.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Crespi B, Nosil P. 2013. Conflictual speciation: species formation via genomic conflict. *Trends Ecol Evol*. 28:48–57.

Dnyansagar VR, Cooper DC. 1960. Development of the seed of *Solanum phureja*. *Am J Bot*. 47:176–186.

Dumbliauskas E, Lechner E, Jaciubek M, Berr A, Pazhouhandeh M, Alioua M, Cognat V, Brukhin V, Koncz C, Grossniklaus U, et al. 2011. The *Arabidopsis* CUL4-DDB1 complex interacts with MSI1 and is required to maintain MEDEA parental imprinting. *EMBO J*. 30:731–743.

Gehring M. 2013. Genomic imprinting: insights from plants. *Annu Rev Genet*. 47:187–208.

Gehring M, Missirian V, Henikoff S. 2011. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* 6:e23687.

Gutierrez-Marcos JF, Pennington PD, Costa LM, Dickinson HG. 2003. Imprinting in the endosperm: a possible role in preventing wide hybridization. *Phil Trans R Soc Lond B* 358:1105–1111.

Haig D. 2013. Kin conflict in seed development: an interdependent but fractious collective. *Annu Rev Cell Dev Biol*. 29:189–211.

Haig D, Westoby M. 1991. Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Phil Trans R Soc Lond B* 333:1–13.

Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, et al. 2011. Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc Natl Acad Sci U S A*. 108:1755–1762.

Jiang H, Köhler C. 2012. Evolution, function, and regulation of genomic imprinting in plant seed development. *J Exp Bot*. 63:4713–4722.

Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecific hybridization. *Curr Biol*. 16:1322–1328.

Jullien PE, Katz A, Oliva M, Ohad N, Berger F. 2006. *Polycomb* group complexes self-regulate imprinting of the *Polycomb* group gene *MEDEA* in *Arabidopsis*. *Curr Biol*. 16:486–492.

Jullien PE, Berger F. 2010. Parental genome dosage imbalance deregulates imprinting in *Arabidopsis*. *PLoS Genet*. 6:e1000885.

Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435–3436.

Köhler C, Page DR, Gagliardini V, Grossniklaus U. 2005. The *Arabidopsis thaliana* MEDEA *Polycomb* group protein controls expression of *PHERES1* by parental imprinting. *Nat Genet*. 37:28–30.

Kradolfer D, Wolff P, Jiang H, Siretskiy A, Köhler C. 2013. An imprinted gene underlies postzygotic reproductive isolation in *Arabidopsis thaliana*. *Dev Cell* 26:525–535.

Lafon-Placette C, Köhler C. 2015. Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Curr Opin Plant Biol*. 23:39–44.

Lester RN, Kang JH. 1998. Embryo and endosperm function and failure in *Solanum* species and hybrids. *Ann Bot*. 82:445–453.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 genome project data processing subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Lin BY. 1984. Ploidy barrier to endosperm development in maize. *Genetics* 107:103–115.

Lu J, Zhang C, Baulcombe DC, Chen ZJ. 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc Natl Acad Sci U S A*. 109:5529–5534.

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A. 2011. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet*. 7:e1002125.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.

Ng DW-K, Lu J, Chen ZJ. 2012. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr Opin Plant Biol*. 15:154–161.

Pacini E, Sarfatti G. 1978. Reproductive calendar of *Lycopersicon peruvianum* Mill. *Bull Soc Bot France* 125:295–299.

Peralta IE, Spooner DM, Knapp S. 2008. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Syst Bot Monogr*. 84:1–186.

Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. 2014. Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife* 3:e03198.

Pires ND, Grossniklaus U. 2014. Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Rep*. 6:63.

Raissig MT, Baroux C, Grossniklaus U. 2011. Regulation and flexibility of genomic imprinting during seed development. *Plant Cell* 23:16–26.

R Development Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. URL: http://www.R–project.org/.

Rick CM. 1979. Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In: Hawkes JG, Lester RN, Skelding AD, editors. The biology and taxonomy of the Solanaceae. New York: Academic Press. p. 667–678.

Rick CM. 1986. Reproductive isolation in the *Lycopersicon peruvianum* complex. In: D'Arcy WG, editor. Solanaceae – biology and systematics. New York: Columbia University Press. p. 477–495.

Rick CM, Lamm R. 1955. Biosystematic studies on the status of *Lycopersicon chilense*. *Am J Bot*. 42:663–675.

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D. 2013. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci U S A*. 110:7934–7939.

Schmid MW, Schmidt A, Klostermeier UC, Barann M, Rosenstiel P, Grossniklaus U. 2012. A powerful method for transcriptional profiling of specific cell types in eukaryotes: laser-assisted microdissection and RNA sequencing. *PLoS One* 7:e29685.

Scott RJ, Spielman M, Bailey J, Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125:3329–3341.

Smeds L, Künstner A. 2011. ConDeTri – A content dependent read trimmer for Illumina data. *PLoS One* 6:e26314.

Städler T, Roselius K, Stephan W. 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. *Evolution* 59:1268–1279.

Städler T, Arunyawat U, Stephan W. 2008. Population genetics of speciation in two closely related wild tomatoes (*Solanum* section *Lycopersicon*). *Genetics* 178:339–350.

Tellier A, Laurent SJY, Lainer H, Pavlidis P, Stephan W. 2011. Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proc Natl Acad Sci U S A*. 108:17052–17057.

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–641.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 28:511–515.

Vu TM, Nakamura M, Calarco JP, Susaki D, Lim PQ, Kinoshita T, Higashiyama T, Martienssen RA, Berger F. 2013. RNA-directed DNA methylation regulates parental genomic imprinting at several loci in *Arabidopsis*. *Development* 140:2953–2960.

Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L. 2009. Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr Biol*. 19:1128–1132.

Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM. 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc Natl Acad Sci U S A*. 110:19639–19644.

Willi Y. 2013. The battle of the sexes over seed size: support for both kinship genomic imprinting and interlocus contest evolution. *Am Nat*. 181:787–798.

Wolf JB. 2009. Cytonuclear interactions can favor the evolution of genomic imprinting. *Evolution* 63:1364–1371.

Wolf JB. 2013. Evolution of genomic imprinting as a coordinator of coadapted gene expression. *Proc Natl Acad Sci U S A*. 110:5085–5090.

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C, Nordborg M, Rehmsmeier M, Köhler C. 2011. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet*. 7:e1002126.

Wolff P, Jiang H, Wang G, Santos-González J, Köhler C. 2015. Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *eLife* 4:e10074.

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U. 2010. *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol*. 20:506–512.

Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y, et al. 2011. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A*. 108:20042–20047.

Figure Legends

FIG. 1. Comparisons of seed set and seed morphology for within-population *versus* interspecific crosses. (*A*) From left to right, the bars quantify the average number of seeds per fruit in crosses among 'sib' plants of LA4329 (*S. chilense*), hybrid crosses with LA1616 plants (*S. peruvianum*) as pollen donors, hybrid crosses with LA4329 plants as pollen donors, and among 'sib' plants of LA1616. White bar proportions correspond to viable seeds, whereas gray proportions indicate shrivelled, empty seeds. Numbers within bars give the number of fruits analyzed per cross type, and error bars indicate standard deviation across fruits. (*B*) Representative seeds for each of the four cross types in (*A*), obtained 60 days after pollination. Note the coiled embryos in normal seeds from within-population crosses (4329sib and 1616sib) and the flat, inviable seeds (aborted embryos) from hybrid crosses, which are much smaller when LA4329 acts as pollen donor. Scale bar, 3 mm.

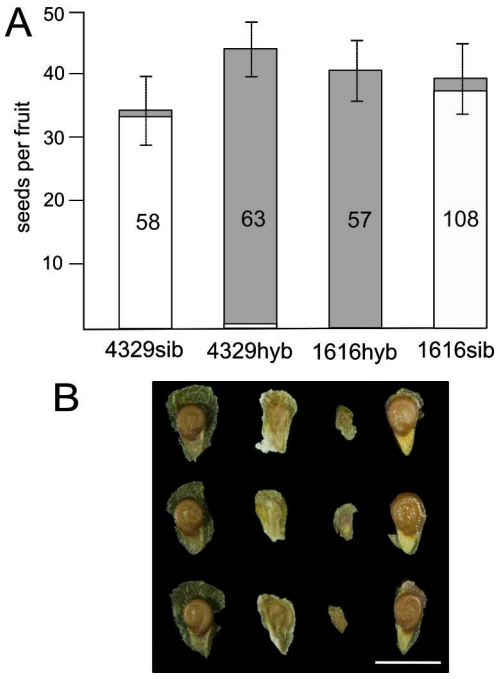
FIG. 2. Global patterns of parent-of-origin-specific maternal proportions in within-population *versus* hybrid endosperm. (*A*) Endosperm maternal proportions for 8,229 genes in the reciprocal *S. peruvianum* crosses 1616A × 1616J (*x* axis) and 1616J × 1616A (*y* axis). Candidate MEGs have maternal proportions >0.833 in both directions of the cross (upper right sector, blue rectangle), and candidate PEGs have maternal proportions <0.333 in both directions of the cross (lower left sector, pink rectangle). Arrow indicates 67 'complete' MEGs (mat. prop. >0.99). (*B*) Endosperm maternal proportions for all 4,111 informative genes in the reciprocal hybrid cross between 1616A (*S. peruvianum*) and 4329B (*S. chilense*). Note the marked shift toward higher maternal proportions in these hybrid endosperms, especially for 1616A in the maternal role (*x* axis, from median maternal proportion 0.646 in *A* to 0.843 in *B*).

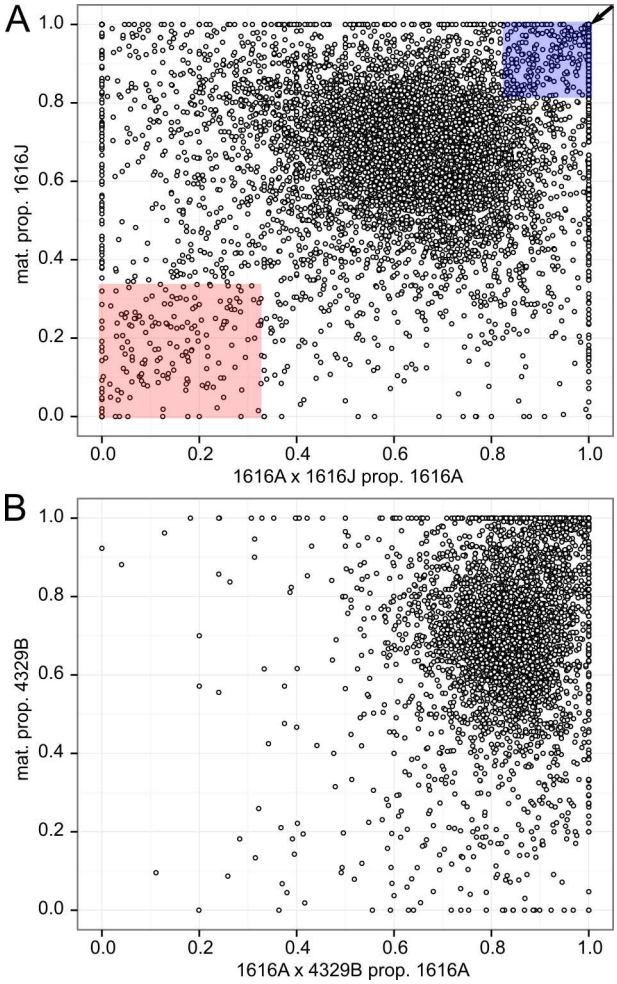
FIG. 3. Global misregulation of maternal proportions in hybrid endosperm and conserved MEGs and PEGs. (*A*) Genome-wide negative correlation between maternal proportion in *S. peruvianum* 1616A sib endosperm (*x* axis) and the difference in maternal proportion between hybrid and sib endosperm (Δ 1616A hyb–1616A sib, *y* axis; Spearman's $\rho = -0.774$, $p < 10^{-15}$). Shown are all 3,647 genes with data on 1616A maternal proportion in both sib and hybrid crosses. The stippled red line at Δ = 0 marks the expectation for genes with no difference in maternal proportion between cross types, and genes above this line show higher maternal proportions in hybrid endosperm. Red dots and blue triangles mark candidate PEGs and MEGs, respectively, whereas other genes in the same ranges of maternal proportions (i.e. <0.333 and >0.833) are not considered as candidate imprinted due to their discordant expression in the 1616J sib endosperm. (*B*) Shift in maternal proportion between within-population and hybrid endosperm for 15 candidate PEGs and 12 candidate MEGs conserved between *S. peruvianum* (LA1616) and *S. chilense* (LA4329). For within-population crosses, PEGs are indicated as red dots and MEGs as blue triangles, and their respective maternal proportion in hybrid endosperm is shown with open symbols. The average shift for PEGs (red arrows along both axes) is less pronounced for *S. chilense* in the maternal role.

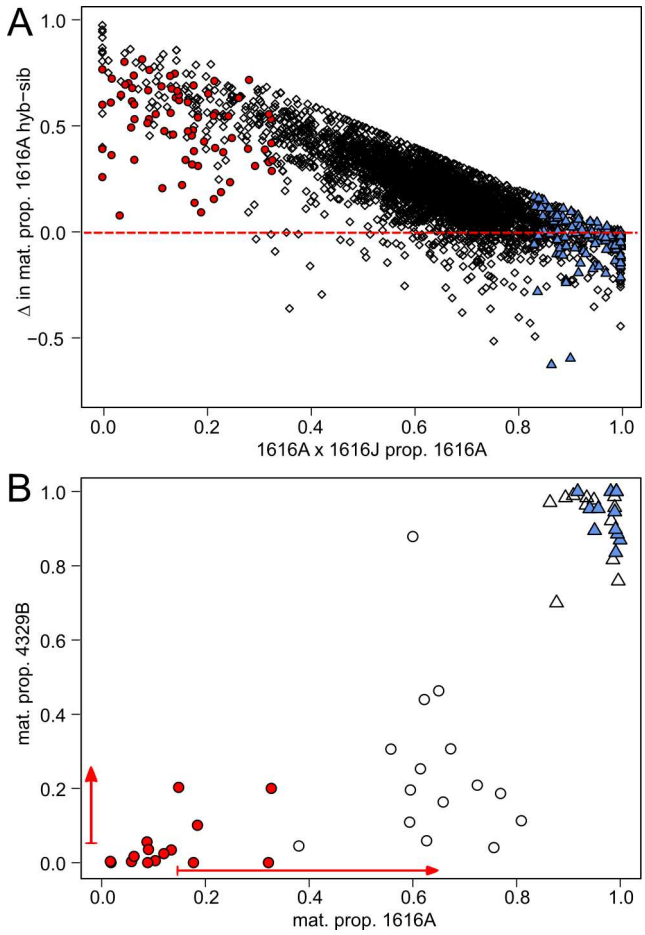
Table 1. Summary of Data Underlying the Estimation of Maternal Proportions in Endosperm-Expressed Genes in Three Reciprocal Cross Types.

Statistic	*peruvianum* cross (1616A ↔ 1616J)	hybrid cross (1616A ↔ 4329B)	*chilense* cross (4329B ↔ 4329K)
Endosperm-expressed genes with SNPs in reciprocal cross	8,654	4,289	2,646
Alternative homozygous sites	6,408	9,225	1,263
Heterozygous sites (e.g. CC:AC)	41,610	13,821	7,421
Mean no. of SNPs per gene	5.55	5.37	3.28
Informative genes after filtering	8,229	4,111	2,560
Candidate MEGs	351	(570)	40
Candidate PEGs	172	(6)	70
Median maternal proportion	0.646 ↔ 0.684	0.843 ↔ 0.701	0.630 ↔ 0.643

NOTE.—Due to the global (asymmetric) misregulation of endosperm expression in the hybrid cross, numbers of candidate MEGs and PEGs for this cross are given in parentheses.


A
seeds per fruit
50
40
30
20
10
58
63
57
108
4329sib
4329hyb
1616hyb
1616sib
B


A
1.0
0.8
0.6
0.4
0.2
0.0
mat. prop. 1616J
0.0
0.2
0.4
0.6
0.8
1.0
1616A x 1616J prop. 1616A
B
1.0
0.8
0.6
0.4
0.2
0.0
mat. prop. 4329B
0.0
0.2
0.4
0.6
0.8
1.0
1616A x 4329B prop. 1616A


A
1.0
0.5
0.0
−0.5
Δ in mat. prop. 1616A hyb−sib
0.0
0.2
0.4
0.6
0.8
1.0
1616A x 1616J prop. 1616A
B
1.0
0.8
0.6
0.4
0.2
0.0
mat. prop. 4329B
0.0
0.2
0.4
0.6
0.8
1.0
mat. prop. 1616A